

Copyright
by
Nayantara Kurpad
2017

**The Report Committee for Nayantara Kurpad
Certifies that this is the approved version of the following report:**

**Investigating Choice and its Relation with Performance, Enjoyment,
Perceived Task Difficulty and Predicted Scores**

**APPROVED BY
SUPERVISING COMMITTEE:**

Supervisor:

Andrew Butler

Diane Schallert

**Investigating Choice and its Relation with Performance, Enjoyment,
Perceived Task Difficulty and Predicted Scores**

by

Nayantara Kurpad, M.S; M.Phil

Report

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Education

The University of Texas at Austin

May 2017

Acknowledgements

Gratitude is small way of expressing my heartfelt thank you to the people who have lived through this research project with me. First, I would like to thank my advisor Dr. Andrew Butler, without whom this research project would not have been possible. He had his doors always open for ideas and suggestions. He has been a constant source of motivation and has steered me in toward the right direction whenever I needed it. I am a better researcher today only because of his training.

I would also like to extend my gratitude to Dr. Diane Schallert for serving as a reader for this report. Her valuable suggestions have been useful to the report. I am grateful to my lab mates – Cynthia Alarcon, Derek Hanson, Lisi Wang and Nathaniel Raley at the ‘Memory Dynamics Lab’ for their valuable feedback. Their suggestions have helped in shaping this report. I am thankful to Megha Joshi, for lending her expertise in quantitative methods. Additionally, I am thankful to Ashwini Ashokkumar, for helping me think through the coding section that this project required.

My experiences in the United States have been enriching only because of my support system. I thank my parents, family and friends for extending their support even though they are miles away. This accomplishment would not have been possible if not for their support. Finally, I thank my husband Arjun Shounak for helping me navigate through graduate school and by providing unconditional love and patience during these long-distance days.

Abstract

Investigating Choice and its Relation with Performance, Enjoyment, Perceived Task Difficulty and Predicted Scores

Nayantara Kurpad, M.Ed

The University of Texas at Austin, 2017

Supervisor: Andrew Butler

Testing situations are often perceived to be negative by students and this affects their final performance on tests. One possible solution to make testing interesting could be by introducing choice during the testing scenario, as choice has shown to increase motivation and interest in learning. Therefore, the present study aims to investigate the role of choice in testing. The study was conducted on 150 participants from Amazon's Mechanical Turk (Mturk). Through random assignment, participants were assigned to one of the five order conditions: easy-hard, hard-easy, random, by block and by choice. Participants were asked to rate their subjective judgements on difficulty, enjoyment and prediction. A one-way ANOVA was conducted and it was found that performance had a significant effect on the order conditions ($F=3.98$, $p < .05$). A Games-Howell post hoc procedure indicated that participants in the random order condition did better than those in the easy-hard condition and the question order by block condition. Further, there was no significant effect of enjoyment, perceived task difficulty and predicted scores on the order conditions. In conclusion, the role of choice in testing was unclear and complex.

Choice was perceived to be difficult instead of being easy. Studies in future can focus on this double-ended nature of choice.

Table of Contents

List of Tables	viii
List of Figures	ix
Chapter 1: Introduction	1
Learning strategies	2
Metacognition in learning	3
Role of choice	4
Present study	6
Chapter 2: Methods	7
Participants	7
Materials	7
Design	8
Procedure	8
Chapter 3: Results	10
Final Performance	10
Perceived Task difficulty	12
Enjoyment	12
Prediction	12
Summary	13
Chapter 4: Discussion	14
Tables	18
Figures	19
Appendix A	23
References	26

List of Tables

Table 1 - Analysis of Variance for Performance	18
Table 2 – Mean and Standard deviation for Performance across conditions.....	18

List of Figures

Figure 1: Box-plots showing performance across the five conditions.....	19
Figure 2: Box-plots showing perceived task difficulties across the five conditions.....	20
Figure 3: Box-plots showing level of enjoyment across different conditions.	21
Figure 4: Box-plots showing the predicted scores across the five conditions.	22

Chapter 1: Introduction

The process of testing has shown to increase learned information and the performance on a test is often related to the learning strategies utilized by students. Learning strategies are the strategies that students engage in while learning new information. Students find it difficult to implement effective learning strategies like retrieval practice and often engage in ineffective learning strategies like rereading. This leads them to have negative perception towards testing. Studies have shown that a flawed learning strategy can have significant impact on performance (Bjork et al., 2013). Although, there is a large body of research has focused on improving student's performance through testing information, fewer studies have paid attention to understanding student's experiences with learning techniques. Student's experiences with learning techniques are related to the underlying cognitive processes that help them draw inferences about their own learning. Their faulty metacognitive judgements often lead them toward ineffective study strategies. Researchers and teachers are constantly trying to understand ways to make learning interesting, and one way to do so could be the provision of choice to students. Provision of choice has shown to increase engagement, interest, effort and persistence on a task (Cordova & Lepper, 1996; Iyengar & Lepper, 1999; Patall et al., 2008, 2010). The present study aims to understand whether introducing choice in testing will increase enjoyment, decrease the perceived task difficulty, and increase performance.

LEARNING STRATEGIES

Numerous learning strategies have been shown to improve memory and performance. Dunlosky et al. (2013) reviewed the evidence for ten learning strategies and on their effectiveness for long-term retention and transfer. The study found that some strategies may not be as effective as some others; for example, strategies like highlighting and rereading were found to have low utility in terms of learning outcomes, while retrieval practice was found to have high utility. Retrieval practice was also shown to promote long-term retention. A study by Karpicke & Blunt (2011) explored this strategy further. They discussed the importance of retrieval and provided evidence for enhanced learning and increased performance when compared to other techniques such as concept mapping and elaborative study techniques.

In spite of evidence leaning toward the use of learning strategies, students tend to use ineffective strategies. Students fall prey to easier and ineffective techniques, like massed learning practices and re-reading because they find it to be easier and more enjoyable. Even though students believe such learning strategies make learning easier, but they actually decrease long-term retention and transfer (Bjork et al., 2013). Kornell & Bjork (2007) showed this in a study where college students were surveyed about their study habit techniques. Only about 18% of the students believed that they learned more when they self-tested compared to rereading whereas about 70% of the students believed that they used self-testing to figure out how well they had learned the information. Therefore, it is possible to conclude that there is an underlying belief that hinders students' use of effective learning techniques. Another reason that could possibly provide support to the use of ineffective strategies is that the difficulty experienced while implementing the learning strategies. The concept of 'desirable difficulties' in learning explains that introducing difficulties in learning would lead to better learning of the

material (Bjork et al., 1994a). Engaging in spaced learning, self-testing are some of the activities known as desirable difficulties (Bjork & Bjork, 2011). Most often, these techniques are harder to implement and impair initial learning but enhance long-term retention. Due to the difficulties in these activities, students tend to give into ineffective strategies like massed study (i.e. rather than spaced study). Hence, it becomes important for learners to understand these difficulties as an opportunity towards better learning.

METACOGNITION IN LEARNING

Underlying the students' selection of learning strategies are cognitive processes which help them make inferences about their own learning. There are several techniques that help explain the process of monitoring their learning. Some of these techniques are judgements of learning (JOLs) (Karpicke & Roediger, 2008), ease-of-learning, confidence ratings etc. These judgements are asked at the end of a complicated task or in between tasks and require complex memory retrieval and monitoring processes. These techniques influence study decisions and maladaptive study decisions are often due to faulty judgements. Additionally, there is evidence that accurate monitoring leads to study choices that enhance learning (Metcalf & Finn 2008; Dunlosky & Rawson 2012). However, studies show that learners do not make accurate evaluations of their learning. Learners are often overconfident and overestimate their learning among other misjudgments (Schneider, 1998). These misevaluations of their learning further lead to lead to poor performance. Similar evidence has been found in studies where students have less confidence in their learning right after the testing process (Karpicke, 2012).

Several factors influence metacognitive judgments. One such factor is fluency, which is the ease and the speed of retrieving information from memory (e.g., Matvey et al. 2001 and Benjamin & Bjork 1996). The judgements involved in judging how well

something is known can sometimes be misleading and can create illusions of knowing. A study by Kelly & Lindsay (1993) provided evidence for participants' confidence on trivia questions. They found that a subject's confidence on an answer was dependent on the time spent in answering the question. Additionally, if they had been given the incorrect answer earlier, they would gain an illusion of knowing. Therefore, fluency is known to influence a lot of different judgements like JOLs, confidence judgements, feeling of knowing etc.

Several studies have looked into ways in which fluency affects judgements. In particular relevance to the current study, a study by Weinstein & Roediger (2010) explored students' performance and subjective evaluations at the end of a quiz. They found that students were had more optimistic evaluations of their performance when they were in the easy-hard condition than the hard-easy condition. A follow up study in 2012, asked students' to evaluate their performance on the last 10 questions they had answered. The results were similar to the previous study which indicated optimistic evaluations on the easy-hard order and while comparable evaluations were seen in the hard-easy condition. However, actual performance did not differ on both the order conditions. Additionally, participants in the easy-hard condition enjoyed the test less and found it more difficult when compared to participants in the hard-easy condition. The conceptual framework from Weinstein & Roediger's (2012) experiments, was used in the present study.

ROLE OF CHOICE

One concept that may positively improve subjective judgements and performance is the provision of choice. Providing choice in the classroom has shown to have motivational benefits. A study by Flowerday & Schraw (2000) found that choice

enhances students' motivation and learning. It also increased engagement with the task and effort on the task. In addition, choice helps to build other skills such as self-regulation. Similarly, other studies have provided positive support for choice in learning environments and that it can lead to increased interest, enjoyment, effort, persistence on a task and perceived competence and task performance (Cordova & Lepper, 1996; Iyengar & Lepper, 1999; Patall et al., 2008, 2010). Additionally, the self-determination theory provides support for choice as one of the determinants that is central to autonomy, motivation and healthy functioning (Ryan & Deci, 2010).

Researchers have identified several factors that influence the effect of choice on learning. These factors include: the number of options or the opportunities for choosing, the cognitive demands of choosing, the background of the participants and the influence of culture, motivation and performance (Iyengar & Lepper, 1999, 2000; Katz & Assor, 2007; Moller et al., 2006; Patall et al., 2008; Patall, 2013; Reeve et al., 2003). For example, a study by Patall (2013) explored interest as a factor that affects choice through three experiments. Results of study 1 showed that individuals reported a range of attitudes (neutral to positive) for making choices. The results also indicated that situations that created greater interest provided positive support for choice. Study 2 indicated that initial interest on the task determined the performance and motivation towards the task. A third study indicated that choice provided motivational benefits when the task was perceived to be boring. On the whole, individuals with high interest seemed to benefit from making choices when compared to individuals with lower levels of interest.

In spite of evidence leaning towards the positive effects of choice, choice can also be detrimental or have no effect on learning and motivation. A number of studies have shown no effect of choice on motivation and performance-related outcomes. In one study, Flowerday & Schraw (2003) found that giving students a choice between working on a

crossword puzzle or an essay task showed no effect on engagement with the task nor did it affect their performance. Additionally, in subsequent studies, they found no effect of choice on a test which assessed students learning. This double-ended nature of choice makes it challenging to understand its impact on learning.

PRESENT STUDY

The primary aim of the present study was to understand the role of choice in influencing students' subjective experiences during testing, as more than often testing is perceived to be difficult. The hypothesis was that choice will increase enjoyment, decrease perceived task difficulty and increase performance on the general knowledge questions. Prior research has shown that choice increases engagement and effort on a task. The present study aimed at understanding the role of choice on the performance on 50 general knowledge questions. The study looked into five question order conditions – easy-hard, hard-easy, random order of questions, questions arranged in blocks and the choice condition. The participants in the choice condition were allowed to make choices based on the category (Science, Animals, Culture, Games and Geography) of their choice. The study also looked into understanding participants subjective judgements like perceived task difficulty, enjoyment and their predicted scores on the quiz.

A secondary aim of the present study was to replicate the results of Weinstein and Roedigers' (2012) experiment. The study found differences in subjective evaluations based on question order conditions. They found that students had more optimistic evaluations of their performance while in the easy-hard condition when compared to the hard-easy condition. The present study implicated the design structure from the experiment to understand the effects of question order.

Chapter 2: Methods

PARTICIPANTS

A total of 150 participants (86 Male, 64 Female) from Amazon's Mechanical Turk (MTurk) participated in the experiment. The target population was restricted to the United States and should have had at least 90% approval rate on previous MTurk HITS (Human Intelligence task). These criteria have been used in prior studies to ensure the quality of data (e.g., Shapiro, Chandler, & Mueller, 2013; Mason & Suri, 2012). Participants were also restricted on their age (18 – 59 years). They were asked for their signed consent at the beginning of the experiment, as approved by the Human Subject Review Board at the University of Texas at Austin. They were compensated a total of \$3 for the completion of the experiment.

MATERIALS

This experiment was conceptualized based on the findings from an experiment conducted by Weinstein and Roediger (2010, 2012). A total of 50 general knowledge questions were selected from the Nelson and Narens' norms (1980) (in Tauber et al., 2013) (See Appendix A). Questions were selected based on the probability of recall of the correct response. The selected questions were within the range of 0.1 to 0.9 difficulty level, where 0.9 indicated easy to answer and 0.1 indicated difficult to answer. Each question had one possible correct answer. For example, "*What animal runs the fastest?*"

and the answer to this question was “*Cheetah*”, “*What is the name of the lightest wood known?*”, answer – “*Balsa*”.

Five categories (science, animals, culture, games and geography were created) and ten questions were chosen from each of these categories. Within each category, the average range of difficulty was around 0.5, with a standard deviation of 0.2. The conditions were balanced such that the average difficulty within the blocks were around 50% according to the original norms.

DESIGN

A between-subjects design was used in this study. Question order was the variable manipulated with five conditions – easy to hard, hard to easy, random order, categorized and category by choice. The dependent variables were performance on the block of 50 questions, ratings of enjoyment and difficulty, and global predictions of overall performance.

PROCEDURE

The experiment was hosted online on Amazon’s Mturk. Participants were randomized into one of the five order conditions. In each order condition, participants were directed to a set of 50 general knowledge questions. Participants were asked to answer all the questions by being honest and if they weren’t sure of an answer they were asked make an educated guess. Participants were asked not to look up answers. They were compensated a total of \$3 at the completion of the survey.

The questions were presented one at a time. After answering all of the 50 questions, participants were asked three metacognitive questions. First, “*How much did you enjoy the test? Please rate your answer by moving the slider*” (none at all – a great deal). Second, “*How difficult did you find this test? Please rate your answer by moving the slider*” (Slightly difficult – extremely easy). Third, “*How many questions do you think you got right? Please enter a number between 0-50*”. At the completion of the metacognitive questions, participants were asked demographic questions like age, sex, ethnicity and race and their Mturk identification.

Participants in the category by choice condition were presented questions based on the category they selected. At first, participants were presented with all the five categories – science, animals, culture, games and geography. They were asked to make a choice and choose one of the five categories. Then, they were presented with four categories which did not include the first chosen category. The process of elimination continued until the participant was left with a single category. After answering questions in each category, participants were directed towards the metacognitive questions and end of survey questions.

Chapter 3: Results

The purpose of the study was to understand the effect of question order on performance, perceived task difficulty, enjoyment and predicted scores. A one-way analysis of variance (ANOVA) was conducted on the five question order condition – whether participants were in the easy-hard condition, hard-easy condition, random order condition, difficulty by block condition and the choice condition as the independent variables.

The following criteria was used to exclude participants – if the participants had reported ‘*don’t know*’ as their answer on five or more than five questions and if they had missing data on the outcomes. Additionally, to determine the outliers – Tukey’s method of outlier detection was used (Hoaglin et al., 1986). This method follows the inter-quartile range of determining the outliers. The first and the third quartile were calculated. After which, the formula to determine the outliers was applied that is $Q_1 - 1.5 * IQR$, where Q_1 is the first quartile (also, the lower limit) and $Q_3 - 1.5 * IQR$, Q_3 refers to the third quartile, the upper limit. Finally, the participants were determined as outliers if they had spent more than 27 minutes on the quiz. In total, 19 participants were excluded from the data set as 9 participants had missing data on the outcomes and 10 participants were outliers.

FINAL PERFORMANCE

The performance on the general knowledge questions were calculated across all the five conditions. The distribution of scores across the five conditions are observed in *Figure 1*. According to the descriptive statistics, it was evident that the average performance differed across the conditions. The random condition had the highest

average performance when compared to the other four conditions ($M=41.67$), while the by block condition had the least average ($M=35.97$).

A prior analysis of the ANOVA assumptions was checked and no influential observations were seen. Levene's test was found to be significant, which violated the homogeneity of variance assumption $F(4,145) = 2.610$, $p\text{-value} < .05$, $\eta^2 = .098$. However, the F ratio is robust because the number of observations in largest group divided by the number of observations in the smallest group is less than the value of 1.5. The independence observation seems reasonable due to random assignment of participants into each condition. The outliers, normality and independence of observations were checked and no problematic observations were found.

The results of the ANOVA is shown in Table 1. The type of question order condition is significant with performance on the test $F(4,149) = 3.95$, $p\text{-value} < .05$, $\eta^2 = .098$. Question order accounted for approximately 9.8% of the variance in performance. In addition, the overall strength of the relationship is low as the partial eta squared of 0.098 is lesser than the cut off of .14 (Cohens, 1977).

Since the groups had unequal group variances, Games-Howell post hoc procedure was used (Table 2). The comparisons indicated that the easy-hard condition and random order condition differed from each other. However, the mean of the random condition was greater than easy-hard condition which means that people in the random condition did better ($M=41.67$, $SD=4.99$). The random condition also differed from the block condition. Participants in the random condition did better than people in the block condition ($M=41.67$, $SD=4.99$).

PERCEIVED TASK DIFFICULTY

Perceived task difficulty was measured at the end of the 50 general knowledge questions. Descriptive statistics indicate that participants in the hard-easy condition found the task to be ‘extremely easy’ ($M=63.37$), when compared to participants in the choice condition ($M=46.20$) who found the task to be ‘slightly easy’ (*Figure 2*).

A one way ANOVA was conducted to understand the mean differences between the different conditions and perceived task difficulty. No significant findings were found $F(4,145) = 1.90, p > .05, \eta^2 = .050$

ENJOYMENT

To understand the participants levels of enjoyment, participants were asked to rate their level of enjoyment toward the end of the 50 general knowledge questions. The means indicate that participants enjoyed the test slightly higher when in the easy-hard condition ($M=82.57$), while compared to the participants in the random, by block and the choice condition as they performed similarly ($M=77$) (*Figure 3*). However, the indicators on the sliders indicate that all the participants enjoyed the test somewhere between ‘a lot’ to ‘a great deal’.

A one way ANOVA was conducted to understand if the group means differed across different conditions. However, no significant findings were found $F(4,145) = .475, p > .05, \eta^2 = .013$

PREDICTION

Participants were asked to report their predicted scores at the end of the test. Descriptive statistics indicate that participants in the random condition predicted their scores higher ($M=41.23$) than participants in the by block condition ended up predicting their scores lesser in comparison ($M=37.20$) (*Figure 4*).

A one way ANOVA was conducted to understand the mean differences in prediction, and no significant findings were found $F(4,145) = 1.097, p > .05, \eta^2 = .029$

SUMMARY

To summarize, only one of measures – performance was found to be significantly different across the five different conditions. These results are in-line with Weinstein and Roedigers' (2012) study, where the participants differed in their performance across the three conditions (easy-hard, hard-easy and random). Post-hoc analysis indicated the easy-hard condition and random order condition differed from each other. Additionally, the random condition also differed from the block condition. The metacognitive measures like perceived task difficulty, enjoyment and predicted scores indicated no group differences among the different conditions.

Chapter 4: Discussion

The primary aim of the present study was to understand the role of choice in testing. Students are often aversive toward tests and have a negative perception towards the testing process. One proposed solution to this problem, was to understand the role of choice in testing because prior research indicated that choice had shown to increase motivation and interest associated with the task (Patall, 2013). Therefore, the present study hypothesized that the provision of choice will lead to increased enjoyment, decrease perceived task difficulty and increase performance on the general knowledge questions. However, the results exhibited the contrary. The results indicated that choice had no significant effect on the outcome measures of performance, perceived task difficulty, enjoyment and predicted scores. Although, certain observable differences were noticed in the choice condition.

The secondary aim of the present study was to replicate the results from Weinstein & Roediger's (2012) experiment. The overall performance across the different conditions in the present study indicated similar results as the earlier experiment. The order conditions were tested and it was found that participants in the random condition performed better than the choice and the hard-easy condition. This finding re-emphasizes the need to order the questions on a test in a random fashion such that participants do not biased to think that they are either doing extremely well on the test or failing the test. Additionally, when the questions were arranged in the hard-easy order, participants believed that they performed better and found the test to be easier. This suggests that a test could be designed in such a way that the easy questions were towards the end in order to help students feel better and therefore exhibit a favorable attitude towards testing.

Comparably, ratings of perceived task difficulty showed similar effects as the experiment by Weinstein & Roediger (2012). In this experiment, participants in the hard-easy condition found the task be easier than participants in the easy-hard condition. In the choice condition, participants were allowed to choose the category of their interest and questions were presented in the order of the categories that they selected. One interesting finding was that participants in the choice condition found the task to be more difficult when compared to the other conditions. There is a good chance that participants picked a certain category because they thought they could relate it to their prior knowledge and experiences. However, towards the end, they were compelled to choose the categories that was not their forte'.

This finding can be possibly explained by a concept known as cue-utilization. This view explains that metacognitive judgements are inferential and are dependent on internal, mnemonic cues that predict or affect ones performance (Koriat & Ma'ayan, 2005). Some of the mnemonic cues are found to be determinants of JOLs, subjective confidence (Schwarz, 2004; Whittlesea & Leboe, 2003).

One such metamnemonic cue that has found to affect subjective experiences is retrieval fluency. Retrieval fluency is the ease of retrieving information in the course of learning (e.g., Benjamin & Bjork, 1996). The participants in the present study utilized the metamnemonic cue of retrieval fluency in responding to the general knowledge questions. Additionally, this is shown to influence their judgement of learning as participants choose the options that they believed to have greater retrieval fluency at the beginning and were forced to choose the difficult option towards the end. This leads them to remember the experiences associated with the difficult questions and therefore conclude by rating the entire test to be difficult.

While the results on the other outcomes like enjoyment and predicted scores did not yield any significant results, certain trends were noticed. The enjoyment ratings were comparable across the easy-hard and hard easy conditions as participants in these conditions rated their enjoyment in a similar fashion. Furthermore, participants in the random condition predicted their performance to be higher than participants in the other four conditions. Although, participants in the choice condition nearly enjoyed the test similarly as participants in the random condition. This finding suggests that there is an undermined role of choice which leads participants to enjoy the test. These findings once again provide evidence towards designing tests with distributed levels of difficulty rather than organizing it from easy-hard or hard-easy.

The role of choice in testing seems to have a complex nature and this make it difficult to infer its affect in testing. The process of making choices involves certain cognitive demands that make it difficult to understand the optimal number of options required to make effective choices. Therefore, it can having a debilitating effect and could also create an illusory effect of autonomy.

The study has certain limitations. First, as the study had five-order conditions, the sample size in the present study was not adequate to detect group differences. As the study was conceptualized based on Weinstein & Roedigers' (2012) experiment, similar results were observed in the present study. Second, the measurement techniques used to measure the metacognitive elements can be revised. A revised version of the measurement scales can reduce the possible variability that was seen in the present study. It could also be possible that the variability could contribute to the lack of significant findings. One possible solution for this is to use likert scales that range from 1-5. In order to control the variability, future research can focus on homogenous samples. Third, information on the time spent on each question could be gathered in order to understand

the participants' retrieval fluency. This would help in understanding if the participants actually struggled to answer the difficult questions.

For future research, the role of choice in testing needs to be investigated further to understand its true effect. As in the present study, choice was offered as categories and the order conditions were not disclosed to the participants. In future, a possible manipulation could be done by disclosing the order conditions to the participants and allowing them to choose between easy-hard, hard-easy and random conditions. Another study can be conducted to understand the role of choice in general knowledge questions versus new learning. Since it is difficult to tease apart the true effect of choice, a stronger manipulation with a larger sample size will be beneficial.

Tables

Table 1 - Analysis of Variance for performance on general knowledge questions.

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	Partial η^2	<i>p</i>
Performance	767.000	4	191.750	3.948	0.098	.005
Error	7041.833	145	48.564			
Total	7808.833	149				

Table 2 – Mean and Standard deviation for performance across conditions.

Condition	<i>n</i>	<i>M</i>	<i>SD</i>
Easy-Hard	50	36.03	6.990
Random order	50	41.67	4.999
Block	50	35.97	8.512

Figures

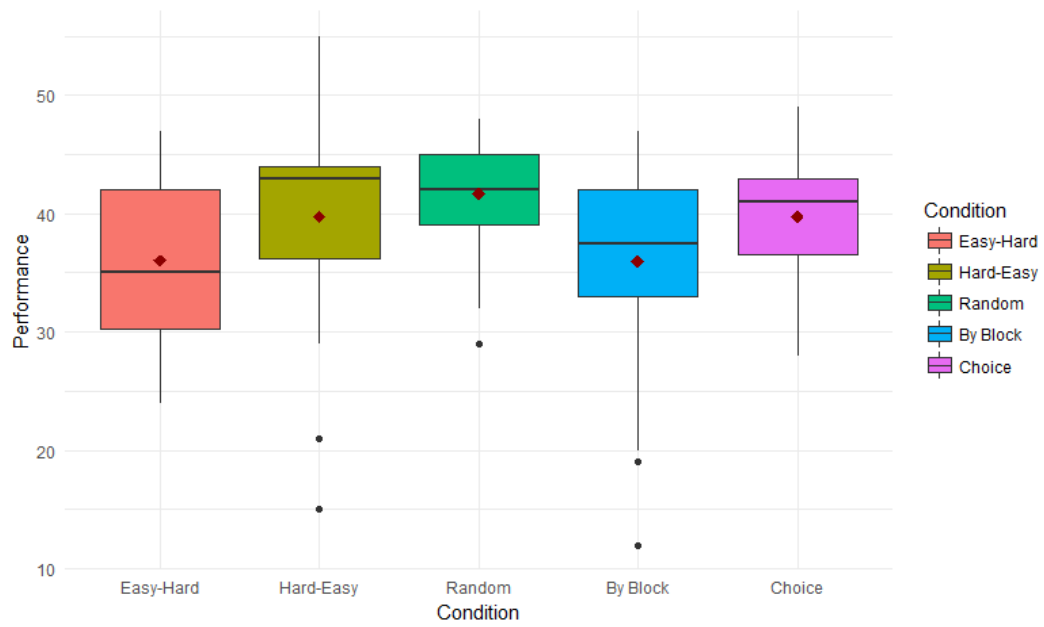


Figure 1: Box-plots showing performance across the five conditions.

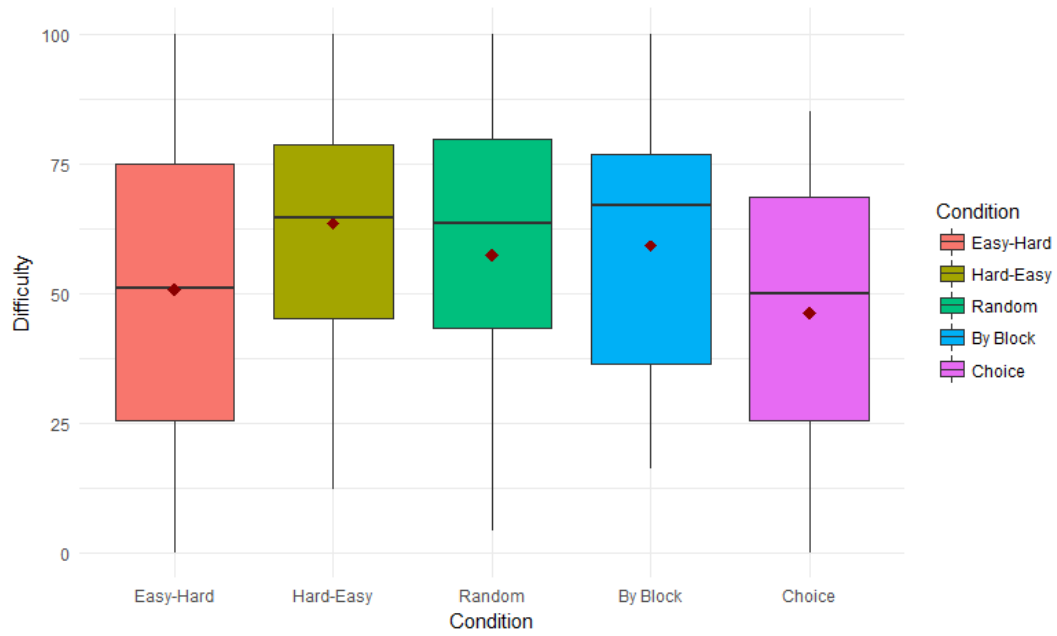


Figure 2: Box-plots showing perceived task difficulties across the five conditions.

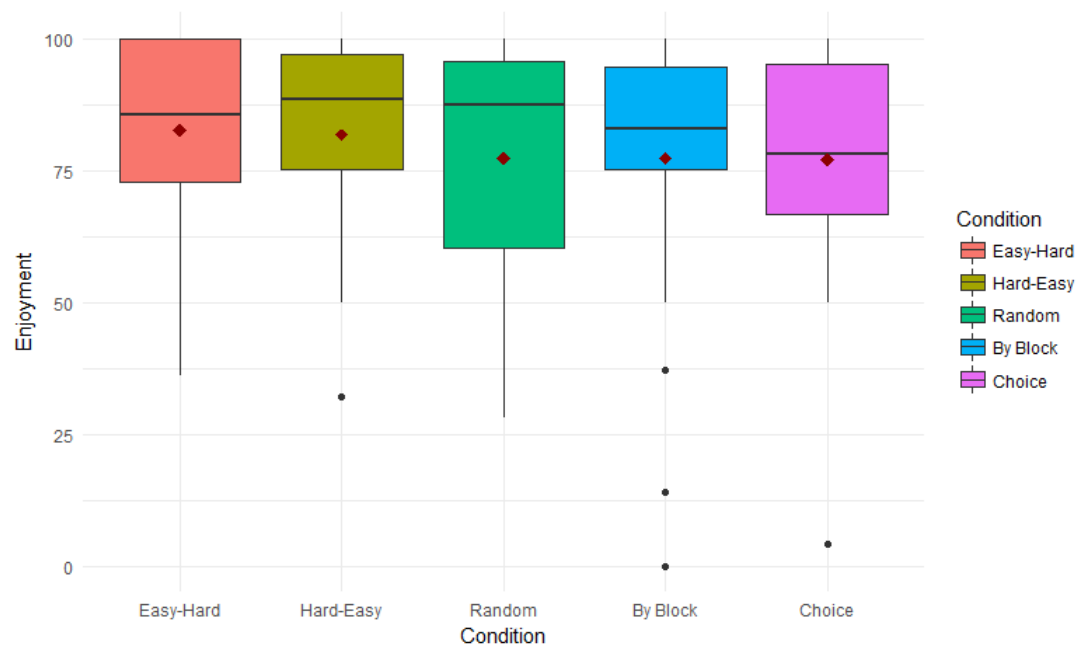


Figure 3: Box-plots showing level of enjoyment across different conditions.

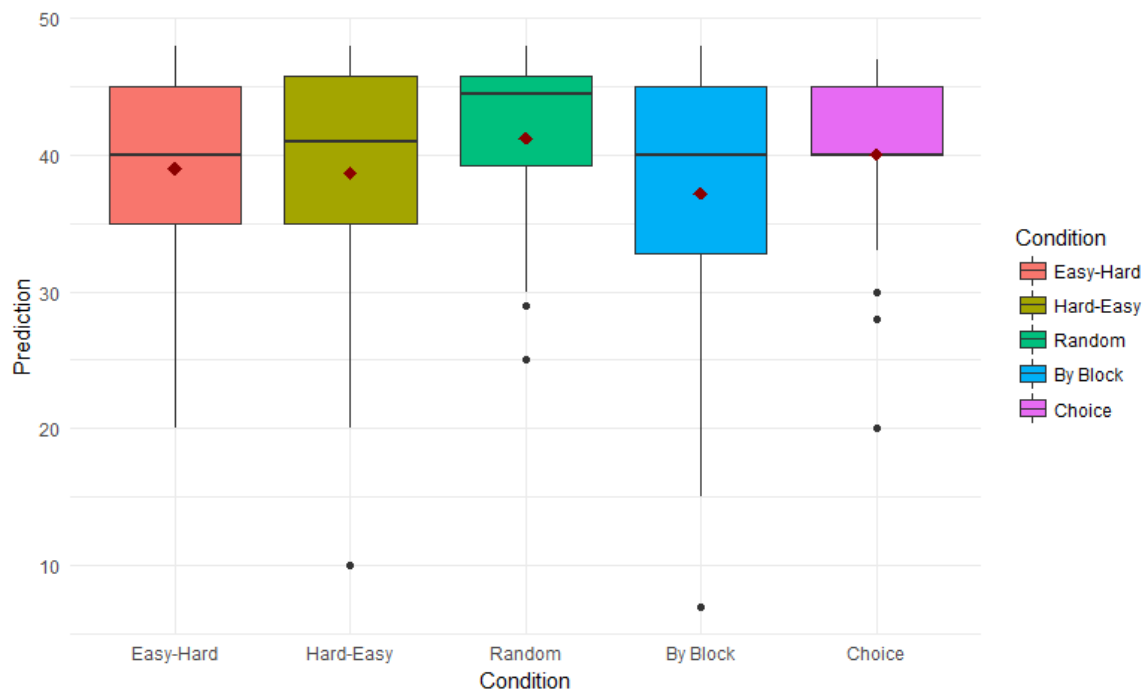


Figure 4: Box-plots showing the predicted scores across the five conditions.

Appendix A

A set of general knowledge questions and their probability of recall.

Questions	Answer	Category	Probability of recall
What is the capital of New York?	Albany	Geography	0.331
What is the name of the city in Italy that is known for its canals?	Venice	Geography	0.459
What is the name of the short pleated skirt worn by men in Scotland?	Kilt	Culture	0.717
Which sport uses the terms gutter and alley?	Bowling	Games	0.788
What is the name of the desert people who wander instead of living in one place?	Nomads	Geography	0.352
What is the name of the three-leaf clover that is the emblem of Ireland?	Shamrock	Culture	0.239
What kind of metal is associated with a 50th wedding anniversary?	Gold	Culture	0.396
What is the name of the lizard that changes its color to match the surroundings?	Chameleon	Animal	0.589
What is the name of the comic strip character who eats spinach to increase his strength?	Popeye	Culture	0.824
What is the name of the organ that produces insulin?	Pancreas	Science	0.327
What is the capital of Russia?	Moscow	Geography	0.309
Which type of snake do Asian snake-charmers use?	Cobra	Animal	0.391
What is the name of the navigation instrument used at sea to plot position relative to the magnetic north pole?	Compass	Geography	0.526
What is the name of deer meat?	Venison	Science	0.432
Which sport is associated with Wimbledon?	Tennis	Games	0.619
Of which country is Baghdad the capital?	Iraq	Geography	0.468
What is the name of the bird that cannot fly and is the largest bird on earth?	Ostrich	Animal	0.603
What is the name of the ocean that is located between Africa and Australia?	Indian	Geography	0.427
What is the name of the furry animal that	Mongoose	Animal	0.149

attacks cobra snakes?			
What is the name of the large hairy spider that lives near bananas?	Tarantula	Animal	0.429
What is the term in golf referring to a score of one under par on a particular hole?	Birdie	Games	0.523
What is the name of the thick layer of fat on a whale?	Blubber	Animal	0.572
What is the name of the severe headache that returns periodically and often is accompanied by nausea?	Migraine	Science	0.847
What is the name of the poker hand in which all of the cards are of the same suit?	Flush	Games	0.486
What is the name of a young sheep?	Lamb	Animal	0.561
What is the name of the spear like object that is thrown during a track meet?	Javelin	Games	0.452
What is the name of the largest ocean on earth?	Pacific	Geography	0.685
What was the name of Tarzans girlfriend?	Jane	Culture	0.781
Which breed of cat has blue eyes?	Siamese	Animal	0.308
In which game are the standard pieces of Staunton design?	Chess	Games	0.138
What is the name for a medical doctor who specializes in cutting the body?	Surgeon	Science	0.748
In what park is old faithful located?	Yellowstone	Geography	0.491
What is the name of batman's butler?	Alfred	Culture	0.163
What is the last name of the famous magician and escape artist who died of appendicitis?	Houdini	Culture	0.436
What animal runs the fastest?	Cheetah	Animal	0.816
In which sport does a rider on horseback hit a ball with his mallet?	Polo	Games	0.51
What is the name of the automobile instrument that measures mileage?	Odometer	Science	0.258
In which sport is the Stanley cup awarded?	Hockey	Games	0.481
What is the name of a dried plum?	Prune	Culture	0.517
What is the name of an airplane without an engine?	Glider	Science	0.157
What is the last name of batman's secret identity in the batman comics?	Wayne	Culture	0.252
What is the last name of the first signer of	Hancock	Culture	0.285

the declaration of independence?			
What is the word that means a nautical mile per hour?	Knot	Science	0.277
What is the capital of France?	Paris	Geography	0.73
What is the name of the lightest wood known?	Balsa	Science	0.125
What is the largest planet in the solar system?	Jupiter	Science	0.559
What is the name for a cyclone that occurs over land?	Tornado	Science	0.641
In which game are men crowned?	Checkers	Games	0.176
What is the term for hitting a volleyball down hard into the opponents court?	Spike	Games	0.784
What is the name of the extinct reptiles known as terrible lizards?	Dinosaurs	Animal	0.315

References

- Bao, X., & Lam, S. (2008). Who makes the choice? Rethinking the role of autonomy and relatedness in Chinese children's motivation. *Child Development*, 79, 269–283.
- Benjamin, A. S., Bjork, R. A. (1996). Retrieval fluency as a metacognitive index. In *Implicit Memory and Metacognition*, ed. LM Reder, pp. 309–38. Mahwah, NJ: Erlbaum
- Bjork, Robert A. Metcalfe, Janet (Ed); Shimamura, Arthur P. (Ed). (1994). Memory and metamemory considerations in the training of human beings. *Metacognition: Knowing about knowing*, (pp. 185-205). Cambridge, MA, US: The MIT Press, xiii, 334 pp.
- Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, 64(1), 417-444. doi:10.1146/annurev-psych-113011-143823
- Bjork, E. L., Bjork R. A (2011). Making things hard on yourself, but in a good way: creating desirable difficulties to enhance learning. In *Psychology and the Real World: Essays Illustrating Fundamental Contributions to Society*, ed. MA Gernsbacher, RW Pew, LM Hough, JR Pomerantz, pp. 56–64. New York: Worth
- Clark, D. A., & Svinicki, M. (2015). The effect of retrieval on post-task enjoyment of studying. *Educational Psychology Review*, 27(1), 51-67. doi:10.1007/s10648-014-9272-4
- Cordova, D. I., & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology*, 88(4), 715-730. doi:10.1037/0022-0663.88.4.715
- Chan, J. C. K., McDermott, K. B., & Roediger, H. L. (2006). Retrieval induced facilitation: Initially nontested material can benefit from prior testing. *Journal of Experimental Psychology: General*, 135, 553–571
- Dunlosky, J, Rawson K.A., (2012) Overconfidence produces underachievement: inaccurate self evaluations undermine students' learning and retention. *Learn. Instr.* 22:271–80

- Dunlosky, J., et al. (2013). "Improving Students' Learning With Effective Learning Techniques: Promising Directions From Cognitive and Educational Psychology." *Psychological Science in the Public Interest*, 14(1): 4-58.
- Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science*, 12, 83–86.
- Flowerday, T., & Schraw, G. (2000). Teacher beliefs about instructional choice: A phenomenological study. *Journal of Educational Psychology*, 92(4), 634-645. doi:10.1037//0022-0663.92.4.634
- Flowerday, T., & Schraw, G. (2003). Effect of choice on cognitive and affective engagement. *The Journal of Educational Research*, 96(4), 207-215. doi:10.1080/00220670309598810
- Hacker, D. J., Bol, L., & Bahbahani, K. (2008). Explaining calibration accuracy in classroom contexts: The effects of incentives, reflection, and explanatory style. *Metacognition and Learning*, 3(2), 101-121. doi:10.1007/s11409-008-9021-5
- Hoaglin, D. C., Iglewicz, B., & Tukey, J. W. (1986). Performance of some resistant rules for outlier labeling. *Journal of the American Statistical Association*, 81(396), 991-999.
- Iyengar, S. S., & Lepper, M. R. (1999). Rethinking the value of choice: A cultural perspective on intrinsic motivation. *Journal of Personality and Social Psychology*, 76(3), 349-366. doi:10.1037/0022-3514.76.3.349
- Kang, S. H. K., Lindsey, R. V., Mozer, M. C., & Pashler, H. (2014). Retrieval practice over the long term: Should spacing be expanding or equal-interval? *Psychonomic Bulletin & Review*, 21(6), 1544-1550. doi:10.3758/s13423-014-0636-z
- Katz, I., & Assor, A. (2007). When choice motivates and when it does not. *Educational Psychology Review*, 19, 429–442
- Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, 319(5865), 966-968. doi:10.1126/science.1152408

- Karpicke, J., & Blunt, J. (2011). Retrieval Practice Produces More Learning than Elaborative Studying with Concept Mapping. *Science*, 331(6018), 772-775. Retrieved from <http://www.jstor.org.ezproxy.lib.utexas.edu/stable/25790300>
- Karpicke, J. D. (2012). Retrieval-based learning: Active retrieval promotes meaningful learning. *Current Directions in Psychological Science*, 21(3), 157-163. doi:10.1177/0963721412443552
- Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language*, 32(1), 1-24. doi:10.1006/jmla.1993.1001
- Koriat, A., & Ma'ayan, H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *Journal of Memory and Language*, 52(4), 478-492. doi:10.1016/j.jml.2005.01.001
- Kornell, N., Rhodes, M. G., Castel, A. D., & Tauber, S. K. (2011). The ease-of-processing heuristic and the stability bias dissociating memory, memory beliefs, and memory judgments. *Psychological Science*, 22, 787-794.
- Marsh, E. J., & Butler, A. C. (2013). Memory in educational settings. *The oxford handbook of cognitive psychology* (pp. 299-317). Oxford;New York;: Oxford University Press.
- Maki, R. H., Berry, S. L. (1984). Metacomprehension of text material. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 663-679. Google Scholar Medline
- Mason, W., & Suri, S. (2012). Conducting behavioral research on amazon's mechanical turk. *Behavior Research Methods*, 44(1), 1.
- Matvey, G., Dunlosky, J., Guttentag, R. (2001). Fluency of retrieval at study affects judgments of learning (JOLs): an analytic or nonanalytic basis for JOLs? *Memory & Cognition*. 29:222-33

- Metcalfe, J., & Finn, B. (2008). Familiarity and retrieval processes in delayed judgments of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 1084–1097.
- Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review*, 15, 174–179. doi:10.3758/PBR.15.1.174.
- Patall, E. A., Cooper, H., & Robinson, J. C. (2008). The effects of choice on intrinsic motivation and related outcomes: A meta-analysis of research findings. *Psychological Bulletin*, 134(2), 270-300. doi:10.1037/0033-2909.134.2.270
- Patall, E. A., Cooper, H., & Wynn, S. R. (2010). The effectiveness and relative importance of choice in the classroom. *Journal of Educational Psychology*, 102(4), 896.
- Patall, E. A. (2013). Constructing motivation through choice, interest, and interestingness. *Journal of Educational Psychology*, 105(2), 504-534. doi:10.1037/a0030307
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55, 68–78
- Reeve, J., Nix, G., & Hamm, D. (2003). Testing models of the experience of self-determination in intrinsic motivation and the conundrum of choice. *Journal of Educational Psychology*, 95, 375–392
- Roderer, T., & Roebers, C. M. (2010). Explicit and implicit confidence judgments and developmental differences in metamemory: An eye-tracking approach. *Metacognition and Learning*, 5(3), 229-250. doi:10.1007/s11409-010-9059-z
- Roediger H., & Karpicke, J. (2006). Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention. *Psychological Science*, 17(3), 249-255. Retrieved from <http://www.jstor.org/stable/40064526>

- Schneider, W. (1998), Performance prediction in young children: Effects of skill, metacognition and wishful thinking. *Developmental Science*, 1: 291–297. doi:10.1111/1467-7687.00044
- Schwarz, N. (2004). Metacognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology*, 14(4), 332-348. doi:10.1207/s15327663jcp1404_2
- Sinkavich F.J. (1995). Performance and metamemory: Do students know what they don't know? *Instructional Psychology*, 22, 77–87. Google Scholar
- Shapiro, D. N., Chandler, J., & Mueller, P. A. (2013). Using mechanical turk to study clinical populations. *Clinical Psychological Science*, 1(2), 213-220. doi:10.1177/2167702612469015
- Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(1), 204-221. doi:10.1037/0278-7393.26.1.204
- Tauber, S. K., Dunlosky, J., Rawson, K. A., Rhodes, M. G., & Sitzman, D. M. (2013). General knowledge norms: Updated and expanded from the nelson and narens (1980) norms. *Behavior Research Methods*, 45(4), 1115-1143. doi:10.3758/s13428-012-0307-9
- Thomas, R. C., et al. (2016). "Prior experience shapes metacognitive judgments at the category level: the role of testing and category difficulty." *Metacognition and Learning* 11(3): 257-274
- Weinstein, Y., & Roediger, H. L. (2010). Retrospective bias in test performance: Providing easy items at the beginning of a test makes students believe they did better on it. *Memory & Cognition*, 38(3), 366-376. doi:10.3758/MC.38.3.366
- Weinstein, Y., & Roediger III, H. L. (2012). The effect of question order on evaluations of test performance: How does the bias evolve? *Memory & Cognition*, 40(5), 727-735. doi:10.3758/s13421-012-0187-3

Whittlesea, B. W. A., & Leboe, J. P. (2003). Two fluency heuristics (and how to tell them apart). *Journal of Memory and Language*, 49(1), 62-79. doi:10.1016/S0749-596X(03)00009-3